VERTICALLY INTEGRATED FLASH EEPROM

FOR GREATER DENSITY AND LOWER COST

5   BACKGROUND OF THE INVENTION

The invention pertains generally to the field of semiconductor, nonvolatile memories, and, more particularly, to the field of vertically-integrated, flash EEROMS which can be manufactured with sufficient density to be cheap enough to compete with rotating magnetic media for bulk memory applications. The vertically-integrated, flash EEPROM according to the

10 teachings of the invention is especially useful in personal computers of the laptop, notebook and palmtop variety although it is broadly applicable to any application where large, nonvolatile memory is needed which is physically rugged and competitive with disk drives in price.

Flash EEPROMS are known in the prior art, but the problem to date has been that they cannot be made cheaply enough for them to have mass market appeal. The size of prior art

15 EEPROM cells has been so large, that the number of cells per semiconductor die that can be made with adequate yield was too low to have a cost which was competitive with rotating memories such as disk drives.

Prior art flash EEPROM cells of the most aggressive design made by Intel Corporation of Santa Clara, California are 7-8 square microns using 0.8 micron design rules. With a

20 semiconductor die size of 1 square centimeter, this cell size allows flash EEPROMS of 4-8 megabits to be built for a cost of about $30 per megabit.

In contrast, small disk drives can be manufactured for about $5 per megabyte. Therefore, a need has arisen for a smaller flash EEPROM cell such that more dense memories can be built for lower cost.

25 SUMMARY OF THE INVENTION

According to the teachings of the invention, a vertically constructed flash EEPROM cell is taught herein which allows a cell size of 2-4 square microns to be achieved. With present 6 inch wafers and 0.8 micron design rules and 40,000-60,000 square mil dies, the cost per megabit of memory cells is a substantial improvement over the $30 per megabit cost of prior

30 art EEPROM cells. With the migration toward 8 inch wafers and 0.6 micron design rules larger die sizes of 100,000-200,000 square mils will be possible, and the cost per megabit of memory cells according to the teachings of the invention should improve greatly.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a cross-sectional view of a semiconductor substrate at an intermediate stage in construction of a vertical EEPROM cell after the recessed gate window has been formed and first polysilicon has been deposited.

Figure 2 is a cross-sectional view of a semiconductor substrate at an intermediat stage in construction of a vertical EEPROM cell after the floating gate has been formed.

Figure 3 is an equivalent circuit showing the two capacitors of the floating gate structure.

Figure 4 is a vertical cross-sectional diagram of a typical prior art EEPROM cell through the floating gate structure.

Figure 5 is a vertical cross-sectional diagram of the finished vertical EEPROM structure.

Figure 6 is a plan view of a cell array using the vertically oriented EEPROM cells according to the invention.

Figures 7A, B and C through Figures 31A, B and C are cross-sectional views showing various stages of simultaneous construction of an NMOS transistor, a PMOS transistor and a vertically oriented EEPROM cell according to a process compatible with fabrication of CMOS drivers for the EEPROM array according to the teachings of the invention.

Figure 32 is a plan view of four cells in an array of EEPROM cells according to the teachigs of the invention.

Figure 33 is a sectional view through a typical EEPROM cell according to the teachings of the invention taken along section line A-A' in Figure 32.

Figure 34 is a sectional view through a typical EEPROM cell according to the teachings of the invention taken along section line B-B' in Figure 32.

DETAILED DESCRIPTION OF THE INVENTION

Referring to Figure 1, there is shown a cross-sectional view of an intermediate stage in the construction of the EEPROM memory cell according to the teachings of the invention. Although a detailed process schedule and series of drawings illustrating the exact method of making one embodiment of the invention will be presented below, Figures 1-3 will be used to summarize the construction of an EEPROM memory cell according to the teachings of the invention.

To reach the stage of construction shown in Figure 1, a one micron deep well is etched into an N type silicon substrate 10 having a resistivity of _____. A P doped region 12

is formed about midway down the well. An N doped region 14 lies above the P type region 12.

An oxide layer 16 having a thickness of about 2000 angstroms is grown on top of the substrate.

An oxide layer 18 is grown at the bottom of the well and has a thickness of about 1000

angstroms. A thin annular oxide layer, sections of which are shown at 20 and 20', is grown on

5 the sidewalls of the well to insulate a first layer of doped polysilicon 22 which is deposited on

the surface of the substrate and into the well.

Figure 2 shows a subsequent stage of construction after an anisotropic etchback to

remove the upper portions of the first polysilicon layer and the first polysilicon lying in the

bottom of the well above oxide layer 18. This leaves a floating gate comprised of an annular

10 first polysilicon layer, two sections of which are shown at 22 and 22'. This floating gate is

isolated from the substrate by the thin oxide layer 20. To complete the electrical isolation of

the floating gate layer 22, a layer of ONO insulator 24 is deposited over the surface of the

substrate and in the well.

The thickness and integrity of the ONO layer is important to the coupling ratio in an

15 EEPROM which is important in the write process. Referring to Figure 3, there is shown an

equivalent circuit of the floating gate and control gate structure shown in Figure 4. Although

Figure 4 represents the structure of a typical prior art floating gate EEPROM structure, it is

used here to illustrate the functioning of an EEPROM cell and the significance to the write

process of the coupling ratio between the capacitance of capacitor C2 and the capacitor C1 in

20 Figure 3. Capacitor C2 represents the capacitor formed between the control gate 31 and the

floating gate 33 in Figure 4. Capacitor C1 represents the capacitor formed between the floating

gate 33 and the substrate 39. Layers 35 and 37 are thin oxide or ONO insulating layers (oxide-

nitride-oxide) that separate the polysilicon one floating gate layer 33 from the substrate 39,

and the polysilicon one floating gate layer from the polysilicon two control gate layer 31,

25 respectively. These two insulation layers separating the conductive polysilicon layers define

the capacitors C1 and C2 in Figure 3. Two oxide spacer layers 51 and 53 insulate the self

aligned edges of the stacked control gate and floating gate structure.

One problem with the prior art stacked structure of Figure 4 was leakage at the corner

57 where ONO is used for insulation layer 37. At this corner, ONO joins the oxide of the spacer

30 layer 51 (the same holds true for the other side) and the electrical seal against charge leaking

out of the floating gate is not perfect because of the concentration of electric field lines at this

corner.

The significance of the coupling ratio pertains to the effectiveness of causing injection of

3

electrons or wells into the floating gate 33 so as to alter the trapped charge therein. It is the presence of trapped charge in the floating gate 33 which alters the threshold of the MOS transistor formed by the floating gate 33, and the source region 41 and the drain region 43 in Figur 4. For one state of trapped charge, an inversion of conductivity type in the substrate 39

5   between the source and drain regions will occur thereby forming a conductive channel through which conduction occurs between the source and drain regions. This channel is symbolized by dashed line 45, and this state of charge can be defined as either a binary 1 or 0. In the other state of charge of the floating gate, no inversion channel occurs, and no conduction between the source and drain occurs. Charge is trapped in the floating gate 33 by tunneling or injection

10 during the write or program process. It is desirable to have the capacitance of capacitor C1 much less than the capacitance of capacitor C2 to insure that sufficient injection or tunnelling of electrons from the source or channel region into the floating gate occurs during the write process. This injection or tunnelling phenomenon occurs when approximatly 15 volts is applied to the control gate terminal 47 in Figure 3 and approximately 8 volts is applied to the

15 source 49 during the write process if C2 is greater than C1. C2 and C1 effectively form a voltage divider between the potential applied to the control gate terminal 47 and the potential of the channel region. It is desirable to have relatively more of the voltage drop from the channel to the control gate terminal 47 occur across capacitor C1 to maximize the tunnelling phenomenon. In other words, when the programming voltage is applied, tunnelling current

20 begins to charge up both capacitors. The smaller capacitor C1 charges up to a higher voltage thereby altering the threshold of the MOS transistor sufficiently to create the inversion channel.

Therefore, since the first oxide layer 35 in Figure 4 or 20 in Figure 2 should be very thin to increase the capacitance of C1 to enhance tunnelling current for writing and erasing, it

25 is necessary for the second oxide layer 37 to be as thin or thinner than the first oxide layer so that C2 is greater than C1. Alternatively, the area of C2 can be made greater than the area of C1. Because of the need for a thin second insulator layer, the material used for the second insulating layer 37 is very important in that it must have high electrical integrity. Generally, ONO is preferred for this purpose because of its high integrity as an electrical insulator and

30 oxide interfaces on both surfaces. Because ONO creates more surface states which would adversely affect the operation of the underlying MOS transistor, ONO cannot be used for the first insulation layer 20 in Figure 2.

ONO layer 24 in Figure 2 is made by oxidizing the underlying layers to a thickness of

4

about 30 angstroms and then depositing approximately 150 angstroms of nitride. Thereafter, steam oxidation of the nitride is performed to form an additional 30 angstroms of oxide. Because of the different dielectric constant of nitride, the overall dielectric constant of the ONO layer 24 is approximately the same as that of 100 angstroms of oxide. ONO works especially well to

5 preserve the trapped charge in the floating gate to alleviate a problem of escaping charge at the corners of the floating gate which existed in the prior art.

After the ONO layer 24 is deposited, a second layer of doped polysilicon 28 is deposited to fill the well and is etched to form the word line.

Figure 5 shows in vertical section the completed device. To reach the state of

10 construction shown in Figure 5, a layer of oxide 29 is grown on the second polysilicon layer 28. Then a mask is formed over the second polysilicon layer 28 to protect the portion thereof overlying the well which it fills. Thereafter, an anisotropic etch is performed to etch down through the polysilicon layer 28, the ONO layer 24, the oxide layer 16 and part of the way through the N-type silicon layer 14 to open a contact well for the bit line 30.

15 After this contact well is opened, an annular oxide spacer, sections of which are shown at 32 and 32', is formed to seal and insulate the sides of the structure from the bit line to be formed next. The oxide spacer is formed by growing or depositing a layer of oxide over the entire structure and anisotropically etching it back to leave the vertical sections of oxide.

The bit line is shared by all devices in a row and is formed by depositing a third layer of

20 polysilicon 30 over the entire structure and etching it to selectively make contact with the N-type silicon layer 14 which forms the drain of the vertical annulus MOS transistor formed inside the well. The source of the vertical MOS transistor is the N-type substrate 10. The channel region for this transistor is formed by the P-type silicon layer 12. The gate oxide between the channel region and the floating gate 22 is oxide layer 20. The control gate is

25 comprised of second polysilicon layer 28, and extends down into the page and up out of the page to form the word line.

Figure 6 shows a plan view of the EEPROM cell. Field oxide 40 defines the outer boundaries of the N-type silicon layer 14 through which the wells 14 and 42 are formed. The polysilicon or metal bit line 30 (polysilicon is shown and preferred for better step coverage)

30 runs from left to right over and in contact with the N-type silicon layer 14 and slightly overlaps the field oxide layer. The bit line also overlaps the word line polysilicon 28 which fills the well 11. The details of the structure down inside the well are not shown in Figure 6 for simplicity.

5

The length of the cell shown in Figure 6 is equal to the dimension A defining the length of the well plus the dimension B which defines the pitch or minimum spacing between the wells. In Figure 6, the next row of wells is represented by wells 48 and 50. For 0.6 micron design rules, A = 0.6 micron and B = 0.6 microns for a total length of 1.2 microns.

5   The width of the cell is equal to the dimension C which defines the width of the well, plus the dimension D which defines the overlap of the second polysilicon layer 28 past the edge of the well, plus the dimension E equal to the pitch between the second polysilicon word lines 28 between columns. For 0.6 micron design rules, C = 0.6 microns, D = 0.05 microns and E = 0.6 microns for a total cell width of 1.3 microns. Thus, the total cell area for 0.6 micron design

10 rules is 1.56 square microns.

With a cell size of 1.56 square microns, a 64 megabit EEPROM memory can be built on a die of 1-2 square centimeter size. With 6 inch wafers, the wafer area is 28 square inches. At 6.54 square centimeters per square inch, a 6 inch wafer contains 182 square centimeters. With a die size of 2 square centimeters, a 6 inch wafer yields about 90 die. Because well known

15 redundancy techniques can be used to repair defective cells, yields in EEPROM production are typically high, averaging around 80 percent. Thus, a typical production run will yield about 72 good die. Typical production costs for a 6 inch wafer are about $500, so the cost per 64 megabit (8 megabytes) die is about $6.94 or about $0.86 per megabyte. A 40 megabyte EEPROM memory using the teachings of the invention would cost about $34.72. This cost should

20 come down with the introduction of 8 inch wafers at 0.6 micron line widths. Typical costs are expected to be about $3.87 per 8 megabyte EEPROM memory or 48 cents per megabyte for a total cost for a 40 megabyte memory of $19.37. Of course any change in any of the numbers of assumptions or numbers used in the above calculations will yield different costs per megabyte. Todays cost for typical prior art EEPROM memory sold by Intel Corporation is about $30 per

25 megabyte manufactured using 0.8 micron design rules. Note that in the above cost calculations, 0.6 micron linewidths were assumed. Costs for prior art EEPROM cells using 0.6 micron design rules should fall to about $15 per megabyte.

A detailed description of how to make the EEPROM memory cell according to the teachings of the invention follows in connection with the discussion of Figures 7A, B and C through Figures

30 30A, B and C. The preferred process is compatible with CMOS processing so that the EEPROM memory can be built on the same die with CMOS drivers. Accordingly, in each of Figures 7A, B and C through Figures 30A, B and C, the figures in the left column labelled Figure _A is the corresponding NMOS structure and the figures in the right column labelled Figure _C is the

corresponding PMOS structure. A summary of the process is given in Appendix A. In Appendix

A, the individual steps in the process are numbered, and the steps in which the masks are used

are given in the column second from the right. The figure numbers in the rightmost column of

Appendix A show the state of construction after the steps preceding the line on which the

5   particular figure number is listed have been completed.

Referring to Figures 7A, B and C, there is shown the state of construction after the first

12 steps in Appendix A. To reach the state of construction shown in Figures 7A, B and C, a P-

type silicon substrate having a conventional resistivity is used as the starting material.

Then a layer of oxide (silicon dioxide) is thermally grown to a thickness of

10   approximately 300 angstroms.

Next a layer of nitride (silicon nitride) is deposited to a thickness of about 1000

angstroms using chemical vapor deposition (CVD), low pressure CVD (LPCVD) or plasma

enhanced chemical vapor deposition (PECVD).

A layer of photoresist is then deposited and developed using the first level twin-well

15   mask to define the twin wells needed to form CMOS devices.

After forming the twin well mask layer of photoresist, the nitride layer is etched away

over an area to be implanted with phosphorous to form an N-type well in which to form the

PMOS device. Any process for etching the nitride will suffice.

To form the N-well, phosphorous is implanted to a depth of about 3000 angstroms using

20   conventional dosage levels. Then the phoshorous is driven in and the N-well area has another

layer of oxide grown thereover using a 1000 degree centigrade oven for one hour. This leaves

the structure as shown in Figure 1 with an N-well 62 for the PMOS device, and N-well 64 in

which the EEPROM device is to be constructed, and a P-well 66 in which the NMOS device is to

be built.

25   Next, the photoresist and nitride are stripped, and boron is implanted to form the P-

well. Both wells are then driven deeper using a 1100 degree centigrade oven for 5 hours to

form wells that art 5-6 microns deep.

The oxide is then etched away over the N-well to clear the substrate surface for further

processing.

30   Finally, a 1000 angstrom nitride layer is grown as shown in Figures 8A, B and C with

the oxide and nitride layers shown as a single layer at 68.

Next, a layer of photoresist is deposited and an active mask (mask 2) is used to cross-

link (develop) sections thereof to leave the structure as shown at Figures 8A, B and C with a

7

photoresist section 70 over the P-well, photoresist section 72 over the EEPROM cell area and

photoresist section 74 over the N-well.

The oxide/nitride layer 68 is then etched using the photoresist as a mask to leave the

structure as shown in Figures 9A, B and C.

5　　　　A field implant must be performed to implant boron at the edges of the active area of the

NMOS device to prevent the formation of parasitic channels, i.e., unintended MOS transistors.

To perform this implant, it is necessary to mask off the N well of the PMOS device. This is done

by depositing a layer of photoresist 76 and developing it with the field implant mask, i.e., mask

3 to leave the second photoresist layer 76 covering the N well. A boron implant is then

10　performed to deposit the P-type field implant impurities shown at 78 in Figure 10A.

After the field implant, the field regions outside the active areas are oxidized to a

thickness of 6000 angstroms to leave the structure as shown in Figures 11A, B and C. The field

oxide is shown at 80. The areas under the field oxide remain doped so they do not invert and

form parasitic MOS devices.

15　　　　Next the fourth mask is used to remove the nitride portion of layer 68 of oxide/nitride

by protecting all structures with photoresist except the oxide/nitride layers 68 over the

EEPROM cells. After developing the photoresist with the fourth mask, a conventional

oxide/nitride etch is performed to leave the structure as shown in Figures 12A, B and C with

photoresist layer 69 protecting the NMOS and PMOS active areas. This leaves a thin layer of pad

20　oxide (not shown) over the EEPROM active areas.

Leaving the photoresist 69 over the NMOS and PMOS wells to protect them, a boron ion

implantion is performed through the pad oxide (not shown) to form the buried P region 82

below the surface of the N well in which the EEPROM cell is to be formed. Typically, the dosage

for this implant will be 1E+12 (on the order of 10 to the 12th power) with an energy level of

25　100 KEV. This implant forms the channel region in the vertical annular EEPROM cell. As the

term annular is used herein, the horizontal cross section through the EEPROM transistor below

the surface of the substrate can be either circular, square, rectangular or some other shape.

Next, leaving the photoresist in place over the NMOS and PMOS devices, an arsenic

implant is performed at a lower energy level to redope the area below the surface of the

30　substrate but above the P region 82 back to N type as shown in Figures 13A, B and C. Typically,

30 KEV is used with a dose of 1E+14.

Still leaving the photoresist in place over the NMOS and PMOS devices, a layer of oxide

84 is grown over the EEPROM cell to leave the structure as shown in Figures 14A, B and C.

The EEPROM cell area will be used to form two vertical EEPROM devices. To start this process, a layer of photoresist is deposited and a fifth mask is used to develop the photoresist so as to open two cell etch windows over the EEPROM cell area. An anisotropic plasma etch process is then used to etch through the oxide layer 86 and etch down into the silicon to form two wells 88 and 90 also called recessed gate windows. These recessed gate windows must hav sufficient depth to penetrate the N layer 86 and the P layer 82 and extend into the N well 64 of the EEPROM cell. This leaves the structure as shown in Figures 15A, B and C.

A pad oxide layer (not shown) 300 angstroms thick is grown next. This layer covers the first nitride layer 68 over the NMOS and PMOS devices, the oxide layer 84 over the EEPROM cells and covers the walls and bottoms of the recessed gate windows 88 and 90. This pad oxide layer protects the underlying structures from a second layer of nitride to be deposited next.

A second layer of nitride 92 approximately 500 angstroms thick is then deposited over the entire structure. This layer covers the walls and the bottom of the two recessed gate windows 88 and 90.

An anisotropic etchback is then performed to remove all portions of nitride layer 92 except those portions on vertical surfaces, i.e., all nitride of layer 92 is removed except those portions on the walls of the recessed gate windows to leave the structure as shown in Figures 17A, B and C.

The nitride of layer 92 is then removed from the walls of the recessed gate windows 88 and 90 using a wet etch to leave the structure as shown in Figures 19A, B and C.

The pad oxide (not shown) underneath the second nitride layer 92 is then removed in a wet etch. Because the pad oxide layer was not separately shown, the structure after its removal looks as shown in Figures 19A, B and C.

A thin gate oxide layer 100 is then grown to insulate the polysilicon floating gate to be formed later from the silicon layers 86 (drain), 82 (channel) and 64 (source). Typically, this gate oxide is grown to a thickness of 90 to 100 angstroms in a process conventional to MOS devices.

Next, a layer of P type doped polysilicon 102 is deposited over the complete structure from which the floating gate 22 in Figure 5 will be formed to leave the structure as shown in Figures 20A, B and C. Typically, about 1000 angstroms of polysilicon is deposited and is doped P type with chemical dope of phosphorous either during or after deposition to a resistivity of 50 ohms per square.

To form the floating gate, the doped polysilicon is etched back off all horizontal surfaces

9

and part way down into the recessed gate windows 88 and 90 to leave the segments of polysilicon shown at 102 in Figure 21B. These segments of doped polysilicon 102 correspond to the floating gate 22 in the finished structure shown in Figure 5.

Electrical isolation of the floating gate is accomplished by formation of another oxide-
5 nitride-oxide layer 104 over the entire wafer to leave the structure as shown in Figures 22A, B, C. Typically, the ONO layer 104 is formed to a thickness of 150 angstroms by a conventional process.

At this point in the process, construction of the NMOS and PMOS devices is started in parallel with the completion of the EEPROM devices. The first step in this process is to deposit
10 a layer of photoresist and develop it with mask 6 to form an ONO protect mask 106 over the EEPROM cell area as shown in Figure 23B. Then an ONO etch and a nitride etch are performed to remove the ONO layer 104 and the nitride layer 68 over the NMOS and PMOS transistor active areas to leave the structure as shown in Figures 23A, B and C. The pad oxide under the nitride layer 68 is left in place to protect the silicon from the threshold adjust implant to be performed
15 next.

A threshold voltage adjustment is next performed by a conventional boron implant to implant charges into the surface region of the N well 62 and the P well 66 to adjust the voltages at which the PMOS and NMOS devices turn on. The design is such that one CMOS device threshold voltage is too low and the other CMOS device threshold voltage is too high before the threshold
20 adjust implant. Then the threshold voltages are adjusted simultaneously in the proper directions by the threshold adjust implant.

After the implant, the pad oxide (not separately shown) that was under the oxide layer 68 is etched away to prepare the NMOS and PMOS devices for growth of a thin gate oxide. During this process the photoresist mask 106 is left in place to protect the EEPROM cell area.
25 A thin gate oxide layer 108 is then grown over the N well 62 and the P well 66 to electrically insulate a gate electrode to be formed later from the underlying silicon. During this process the photoresist mask 106 is left in place to protect the EEPROM cell area.

Next, the photoresist mask 106 is removed, and a second doped polysilicon layer 110 is then deposited to a thickness of about 3000 angstroms. The control gates for the PMOS, NMOS
30 and EEPROM devices will be formed from this polysilicon layer 110. This second polysilicon layer also fills the recessed gate windows 88 and 90 and covers the ONO layer 104.

A thin layer of silicon dioxide 112 is then grown over the entire second polysilicon layer 110 to a depth of about 2000 angstroms.

10

A seventh mask is then used to develop a layer of photoresist deposited over the second polysilicon layer 110 and oxide 112 for purposes of etching the second polysilicon layer to form the control gates of the PMOS and NMOS devices and of the EEPROM cells and the word lin s corresponding to word line 28 in Figures 5 and 6. After the etch of the second polysilicon by a conventional process, the structure looks as shown in Figures 25A, B and C xcept that an LDD phosphorous implant to form the source and drain regions of the NMOS device has not yet been performed.

To form the source and drain regions of the NMOS devices, an 8th mask is used to develop a layer of photoresist to form an LDD implant mask over the PMOS and EEPROM devices. Then phosphorous is implanted in a conventional process using the etched second polysilicon layer 110 over the NMOS device as a mask to form self aligned LDD regions (lightly doped drain regions) shown at 114 in Figure 25A. Later, more heavily doped, deep source and drain regions will be formed, but the LDD implants prevent short channel problems.

To protect the sidewalls of the control gates of the NMOS, PMOS and EEPROM devices, a spacer oxide deposition is performed to a depth of 3000 angstroms and then the spacer oxide is etched back to form the spacer oxide regions 114 on the sidewalls of the polysilicon control gates formed from second polysilicon layer 110. The spacer etch is an anisotropic etch to remove the spacer oxide from only the horizontal surfaces.

Referring to Figures 27A, B and C, to open contact holes 118 and 120 to the EEPROM cell, a layer of photoresist is deposited and developed with a ninth mask to form a cell contact etch mask layer 116 protecting the PMOS and NMOS devices. The developed photoresist of layer 116 is also located so as to bound the outer limits of the contact holes to be etched through the ONO layer 104 and the oxide layer 84. The other boundaries of these contact holes are self aligned with the outer edges of the spacer oxide 114. Oxide layers 113 are then formed on top of the second polysilicon control gates 110 using the photoresist 116 as a mask as shown in Figure 28B.

The ONO etch and oxide etch is then performed to leave the structure as shown in Figures 27A, B and C with contact holes 118 and 120 to the N type layer 86 for the bit line connections (not shown).

To form the bit lines corresponding to the bit line 30 in Figures 5 and 6, a layer of metal or polysilicon 122 is deposited over the structure. Metal is shown at 122 in Figure 28B, but doped polysilicon is preferred for better step coverage.

11

Photoresist is then deposited and a tenth mask is used to develop it to form a protective layer over the EEPROM devices so as to allow removal of the metal or polysilicon off the NMOS and PMOS devices and so as to define the outlines of the bit lines. The metal or polysilicon 122 is then etched into the shape of the bit lines and removed from over the PMOS and NMOS devices

5 to leave the structure as shown in Figures 28A, B and C.

Next, to complete the NMOS device, an N+ arsenic implant must be performed in the P well. To accomplish this, a layer of photoresist is deposited and developed with an eleventh mask to protect the EEPROM cell and the PMOS active area by photoresist which is not shown in the figures. An N+ arsenic implant is then performed using this photoresist exposing the P well

10 and the polysilicon 110 and the spacer oxide 114 as a mask to form the self-aligned source and drain regions 130 and 132.

To complete the PMOS device, another layer of photoresist is deposited and developed with mask 12 to expose the N well 62 and protect the EEPROM active area and the P well 66. A P+ boron implant is then performed using this photoresist as a mask and the second polysilicon

15 control gate 110 and spacer oxide 114 as a mask to form self-aligned source and drain regions 134 and 136. This leaves the structure as shown in Figures 29A, B and C.

To repair the implant damage, the structure is annealed 1000 centigrade for 30 seconds.

To passivate the structure, a BPSG deposition is performed to a thickness of 6000 angstroms.

20 To complete the NMOS and PMOS devices, contacts to the source and drains of the PMOS and NMOS devices must be made. To do this, a layer of photoresist is deposited and developed using contact mask 13. An etch is then performed to cut the contact holes 138, 140, 142 and 144 through the BPSG layer 146.

After a contact reflow to soften the edges for better step coverage , a layer of metal is

25 then deposited to 7000 angstroms and etched to form the contacts 148, 150, 152 and 154 to complete the structure as shown in Figures 31A, B and C.

Referring to Figure 32, there is shown a plan view of four cells in an array of vertically oriented EEPROM cells according to the teachings of the invention and constructed according to a process which is compatible with the simultaneous formation of CMOS devices on the same die.

30 The outlines of two recessed gate windows in which two EEPROM cells are formed are shown at 88 and 90. First polysilicon word lines are shown at 110. The metal or second polysilicon bit lines are shown at 122. The drain regions of the EEPROM cells are shown at 123 and 125.

Figure 33 is a cross-sectional view taken along section line A-A' in Figure 32 of the

lower two EEPROM cells having recessed gate windows shown at 127 and 129 in Figure 32.

Figure 34 is a cross-sectional view of the EEPROM cells in recessed gate windows 90 and 129 in Figure 32 taken along section line B-B' therein. Structural elements in Figures 33 and 34 corresponding to elements in Figures 7A, B and C through 31A, B and C and Figure 32 have the same reference numerals.

There is given below a table summarizing the above described process of building the flash EEPROM according to the teachings of the invention which is compatible with simultaneous fabrication of CMOS devices on the same die.

PROCESS FLOW FOR CONSTRUCTING A SELF-ALIGNED EEPROM MEMORY CELL COMPATIBLE WITH CMOS DRIVERS ON THE SAME DIE

STEP	DETAILS	MASK	FIGURE
1. Start with silicon substrate	P-Type, Resistivity __		
2. Grow a layer of oxide	Approx. 300 angstroms		
3. Deposit a layer of nitride	Approx. 1000 angstroms		
4. Deposit and develop a layer of photoresist using twin well mask		Mask 1	
5. Etch nitride layer over portion of substrate to become N-well 62			
6. Form N-well 62 with phosphorous implant	3000 anstroms deep, conventional dosage		
7. Drive phosphorous and re-oxidize N-well 62	1000 degrees C, 1 hour		
8. Strip photoresist and nitride			
9. Implant Boron to form P-well 66			
10. Drive the N and P wells 62, 64 and 66 deeper	1100 degrees C, 5 hours, 5-6 microns deep after drive		Figure7A, B and C
11. Etch oxide over N-well 62 to clear the surface thereof for further processing			
12. Grow pad oxide	300 angstroms		
13. Deposit nitride layer	1000 angstroms		
14. Deposit photoresist and use active mask to develop photoresist to define etch masks 70, 72, 74 for active areas		Mask 2	Figure 8A, B and C
15. Etch oxide/nitride layer 68 to define active areas			Figure 9A, B and C
16. Deposit a layer of photoresist and develop using a field implant mask to form field implant mask 76		Mask 3	Figure 10A, B and C
17. Boron implant to deposit field implant impurities in P well.	Conventional dosage and energy		
18. Grow field oxide	6000 angstroms		Figure 11A, B and C
19. Deposit photoresist and develop with mask 4 to leave exposed only the ONO layer 68 over the EEPROM cells		Mask 4	
20. Etch away ONO layer 68 over EEPROM cell			Figure 12A, B and C

21. Implant boron to form P region 82 below substrate surface throughout N well in which EEPROM is to be formed to make channel region 12 of finished device as shown in Figure 5.	100 KEV, 1E+12		
22. Implant arsenic to redope to N type region 86 below surface but above P layer 82	30 KEV, 1E+14		Figure 13A, B and C
23. Grow layer of oxide 84 over EEPROM cell area	2000 angstroms		Figure 14A, B and C
24. Deposit layer of photoresist and use cell etch mask 5 to develop to open windows for etching recessed gate windows 88 and 90		Mask 5	
25. Anisotropically etch recessed gate windows 88 and 90 through N layer 86 and P layer 82 into N well 64			Figure 15A, B and C
26. Grow pad oxide layer over whole substrate to protect underlying structures from second nitride layer	300 angstroms		
27. Deposit second nitride layer 92 which is thinner than first nitride layer 68	500 angstroms		Figure 16A, B and C
28. Perform nitride etchback to remove nitride of layer 92 on all horizontal surfaces and leave it covering only the vertical walls of the recessed gate windows 88 and 90	anisotropic etch		Figure 17A, B and C
29. Grow oxide 96 on bottoms of recessed gate windows	2000 angstroms		Figure 18A, B and C
30. Cell nitride strip	dip off nitride in wet etch		Figure 19A, B and C
31. Pad oxide strip	dip off pad oxide in wet etch		
32. Grow thin gate oxide layer 100	90-100 angstroms, conventional process		
33. Deposit doped polysilicon layer 102 from which floating gate is to be formed	1000 angstroms doped P type to 50 ohms per square		Figure 20A, B and C
34. Etch back doped polysilicon layer 102 from horizontal surfaces to leave floating gates			Figure 21A, B and C
35. Form Oxide-Nitride-Oxide layer 104 above floating gates	Conventional process, 150 angstroms		Figure 22A, B and C
36. Form ONO protect mask 106		Mask 6	
37. ONO etch, nitride etch to clear PMOS and NMOS active areas for transistor formation			Figure 23A, B and C
38. Threshold voltage adjust implant	Boron		
39. Leaving photoresist mask 106 in place, etch away pad oxide under first nitride layer 68 to expose N well and P well silicon			
40. Leaving photoresist mask 106 in place, grow thin gate oxide 108 over N well 62 and P well 66	150 angstroms		
41. Remove photoresist mask 106, and deposit doped second polysilicon layer 110 over entire structure	3000 angstroms		
42. Oxidize second polysilicon	2000 angstroms		Figure 24A, B and C

43. Deposit photoresist, and use 7th mask to develop a second poly etch mask		Mask 7	
44. Etch second polysilicon 110 and overlying oxide to form control gates and word lines corresponding to word line 28 in finished device of Figure 5			
45. Deposit photoresist and develop using 8th mask to protect PMOS and EEPROM devices to form LDD implant mask		Mask 8	
46. Phosphorous LDD implant using control gate poly as a mask to form self-aligned LDD regions of NMOS devices.	Conventional process		Figure 25A, B and C
47. Deposit spacer oxide	3000 angstroms		
48. Anisotropically etch spacer oxide to leave spacers on sidewalls of polysilicon control gates.			Figure 26A, B and C
49. Deposit photoresist and develop with Mask 9 to protect the NMOS and PMOS devices for a bit line contact hole etch and reoxidize tops of second polysilicon 110 to form oxide layer 113		Mask 9	Figure 27A, B and C
50. Etch self aligned bit line contact holes 118 and 120 through ONO 104 and oxide 84			
51. Deposit bit line metal or poly 122	5000 angstroms		
52. Deposit layer of photoresist and develop using 10th mask to form protective mask layer over NMOS and PMOS devices		Mask 10	
53. Etch bit line metal layer 122 to form bit lines			
53. Deposit photoresist and develop using mask 11 to expose P well 66 and protect EEPROM active area 64 and N well 62		Mask 11	
54. N+ arsenic source and drain implant for NMOS device	Arsenic, conventional process	Mask 12	Figure 29A, B and C
55. Anneal implants	1000 C, 30 sec		
56. BPSG passivation deposition	6000 angstroms		
57. Deposit photoresist and develop with contact mask 13 to form mask for contact holes for NMOS and PMOS devices		Mask 13	
58. Etch contact holes			
59. Contact reflow		Mask 14	
60. Metal deposition, mask and etch to form contacts 148, 150, 152 and 154	7000 angstroms	Mask 15	Figure 31A, B and C

Although the invention has been disclosed in terms of the preferred and alternative embodiments described herein, those skilled in the art will appreciate different variations and alternatives which may be used to embody the teachings of the invention. All such variations and alternatives are intended to be included within the scope of the claims appended hereto.